



Quantifying Proficiency Gains: Evaluating the Efficacy of a One-Month TOEIC Preparation Program

Lester L. Potot¹, Dennishe S. Tumulak², Jhonwindel B. Tibalan³, Erwin S. Quiamco⁴

¹Center For English Language Studies, Lapulapu-Cebu International College, Philippines

Email: lester.potot@cels.lcic.edu.ph

²Center For English Language Studies, Lapulapu-Cebu International College, Philippines

Email : dennnytumulak@gmail.com

³Center For English Language Studies, Lapulapu-Cebu International College, Philippines

Email: tibalanjh@gmail.com

⁴Center For English Language Studies, Lapulapu-Cebu International College, Philippines

Email :erwin.quiamco@cels.lcic.edu.ph

Received: 20 April 2026; Received in revised form: 22 May 2026; Accepted: 26 Jun 2026; Available online: 29 Jun 2026

©2026 The Author(s). Published by Infogain Publication. This is an open-access article under the CC BY license

(<https://creativecommons.org/licenses/by/4.0/>).

Abstract—This study examined the comparative test of English for International Communication (TOEIC) Reading and Listening (RL) performances of 194 English as a Foreign Language (EFL) students from Taiwan enrolled at the Center for English Language Studies (CELS) of Lapulapu-Cebu International College during the 2024–2025 academic year. Using a descriptive quantitative research design, the study analyzed students' pretest and posttest TOEIC scores to assess the efficacy of an intense English immersion program lasting one month. Four to five focused language modules that provided 24 hours of in-person training per course made up this pedagogical intervention. The 90-minute daily sessions operationalized structured language exposure using a Task-Based Language Teaching (TBLT) framework that balanced performance-driven communicative output activities specific to standardized test domains with real-world listening and reading input. Results revealed a notable improvement in students' overall proficiency, with the mean score increasing from 423.25 ($SD = 152.57$) in the pretest to 503.89 ($SD = 173.25$) in the posttest. A paired sample t -test indicated a statistically significant difference between the two means ($t = 11.93, p < .001$), confirming the positive impact of the intervention. Furthermore, 85% of participants demonstrated score gains, and 48 students advanced to higher CEFR levels. The findings suggest that focused instruction, task-based learning, and consistent TOEIC-oriented practice significantly enhance learners' communicative competence. This study underscores the effectiveness of short-term intensive English programs in improving EFL students' test performance and overall language proficiency.



Keywords— TOEIC RL, EFL learners, language proficiency, test performance, English instruction, CELS, CEFR level

I. INTRODUCTION

In an increasingly interconnected world, English proficiency has become a fundamental requirement for academic success and professional advancement. As English continues to dominate international communication, business, and education, the ability to use the language effectively is considered a key factor for global competitiveness (Crystal, 2012). To assess English

proficiency, standardized language tests such as the Test of English for International Communication, Reading and Listening (TOEIC RL) have gained significant importance, particularly among non-native speakers seeking educational and career opportunities abroad (Educational Testing Service [ETS], 2023).

The TOEIC test, developed by ETS, is designed to measure the English listening, reading, speaking, and writing

abilities required in real-world workplace contexts (Powers & Kim, 2019). It serves as an objective benchmark of communicative competence, allowing institutions and employers to evaluate individuals' readiness for participation in international settings. For foreign students, especially those studying or planning to work in English-speaking environments, TOEIC scores often determine eligibility for admission, scholarships, internships, or employment opportunities (Lee, 2020). Hence, TOEIC performance not only reflects linguistic ability but also influences students' academic trajectories and future career prospects.

Moreover, the process of preparing for the TOEIC test may foster broader language learning outcomes. Research suggests that test preparation enhances learners' motivation, promotes strategic learning behaviors, and improves their overall communicative competence (Choi, 2021; Green, 2014). Consequently, the TOEIC RL test can be seen not only as an evaluative tool but also as a catalyst for language development and intercultural competence among foreign students.

Given its growing role in international education and employment, it is essential to examine the significance of the TOEIC RL test among foreign students who studied in the Philippines. This study aimed to present the test performances of the international students of Lapulapu-Cebu International College (LCIC) and draw implications on the influence of studying in Cebu particularly in the Center for English Language Studies (CELS) as reflected in their pre test and posttest achievements.

Research Objectives

This study evaluated the pedagogical efficacy of a short-term intensive English immersion curriculum—specifically looking at the Quantifying Proficiency Gains: Evaluating the Efficacy of a One-Month TOEIC Preparation Program—among Taiwanese EFL learners enrolled at the Center for English Language Studies (CELS) of Lapulapu-Cebu International College during the 2024–2025 academic year. The empirical findings from this investigation served as the objective foundation for a strategic institutional action plan.

Specifically, the study sought to answer the following questions:

1. What are the pretest and posttest performances of the experimental group?
2. Is there a significant difference between the pretest and posttest performances of the experimental group?
3. What changes are observed in the students' CEFR proficiency levels following the intensive program

4. Based on the results of the study, what implications can be drawn and an action plan can be proposed?

II. THEORETICAL BACKGROUND

This study's theoretical framework is intended to assess the pedagogical effectiveness of the Center for English Language Studies' (CELS) short-term concentrated Test of English for International Communication Reading and Listening (TOEIC RL) preparation program. The study blends two complimentary supporting cognitive frameworks and anchors itself inside one fundamental macro-theory of second language learning in order to rigorously explain the linguistic changes seen throughout the student cohort. When taken as a whole, these interconnected models create a coherent causal chain that explains how high-stakes linguistic output demands, strategic structural scheduling, and intense task engagement work together to accelerate global language proficiency transitions within a thirty-day timeframe.

The overarching anchor theory governing the instructional delivery of the intervention is Michael Long's operationalization of Task-Based Language Teaching (TBLT), which extends from the foundational tenets of the Interaction Hypothesis (Long, 1996, 2015). Long maintains that second language learning does not occur through the passive, rote memorization of structural grammar principles or isolated vocabulary lists typical of traditional foreign language schools. Instead, linguistic development is dramatically accelerated when adult learners are integrated within authentic, goal-oriented communication situations where they must actively negotiate for meaning (Long, 2015). When students encounter communication breakdowns or understanding hurdles during a targeted communicative activity, they are cognitively pushed to "notice the gap" between their current interlanguage state and the correct target structure (Schmidt, 2001).

While TBLT provides the interactive structural container for classroom delivery, Merrill Swain's (1985, 2005) Pushed Output Hypothesis serves as the first supporting theory to explain the precise mechanism behind the dramatic proficiency gains achieved by the students. In particular, the study's empirical data showed that a significant percentage of the students who were progressing were able to overcome the persistent "A2 proficiency plateau" and effectively cross the independent B1 criterion. Because the environment lacks the strain necessary to promote active grammatical restructuring, lower-intermediate foreign language learners often gain high levels of passive input comprehension but stay frozen at a basic user level, according to Second Language Acquisition research (Richards, 2015).

In order to close this gap, Swain's approach proposes that learners can only achieve real linguistic fluency and structural revision when they are methodically "pushed" by their learning environment to produce coherent, accurate, and contextually appropriate language (Swain, 2005). Three crucial metacognitive functions are served by this pushed output, according to Swain: a noticing function where learners identify their internal production limitations; a hypothesis-testing function where they try out novel syntactic forms; and a metalinguistic function where they deliberately consider structural accuracy (Swain, 1995, 2005).

By allocating half of the institutional grading weight to a continuous track of weekly Mini-Tests and rigorous Unit Assessments, the CELS curriculum directly initiates these three functions. The classroom becomes an active production crucible instead of a passive listening lecture thanks to this rigorous testing atmosphere, which serves as the exact catalyst needed by the Pushed Output Hypothesis. Students are forced to perform sophisticated syntactic processing—analyzing precise word endings, verbal inflections, and passage coherence in Parts 5, 6, and 7 of the reading syllabus—instead of basic semantic guessing due to the temporal and structural demands of these evaluations. The lower-tier pupils are eventually able to overcome the false-beginner ceiling and become self-regulated, independent language users thanks to this persistent effort.

Lastly, the study uses the Interleaving Effect, which was developed by educational psychologists Richard Schmidt and Robert Bjork (1992) and supported by John Sweller's (1988) Cognitive Load Theory, as its second supporting framework to explain the distinct statistical distribution found in the post-test results. A crucial distinction in educational design is made between "interleaved" practice schedules, in which several distinct skills are methodically mixed and alternated within a brief period of time, and "blocked" practice schedules, in which a single skill is isolated and fully practiced before going on to the next topic (Schmidt & Bjork, 1992). While sequential, blocked schedules frequently create the appearance of quick mastery during immediate classroom exercises, a wealth of cognitive research shows that blocking often results in catastrophic cognitive fatigue and long-term skill attrition during real-world standardized testing scenarios where quick attentional switching is required (Dunlosky et al., 2013).

III. METHODOLOGY

Research Design

This study utilized a quantitative approach using a quasi-experimental, single-group pretest-posttest design. This design was selected to measure the systematic effects of the one-month intensive pedagogical intervention on one group where true random assignment was not structurally feasible due to institutional grouping. Descriptive quantitative techniques were supplementary implemented to map out frequencies, percentages, and baseline profile distributions across CEFR bands.

Research Participants

All 194 respondents are university students from Taiwan, specifically National Kaohsiung University of Science and Technology and Chaoyang University of Technology. The participants completed a pretest in TOEIC RL as administered in their respective schools in Taiwan as their pretest performance before their arrival in the Philippines and were placed into specific CEFR levels.

Research Instrument

This study used the official, standardized institutional Test of English for International Communication (TOEIC) Reading and Listening (RL) examination created by the Educational Testing Service (ETS) as the main assessment of improvements in language proficiency. A 100-item Listening Comprehension section (Parts 1 to 4) lasting roughly 45 minutes and a 100-item Reading Comprehension section (Parts 5 to 7) lasting 75 minutes make up the instrument's two main components, which are administered over a strict two-hour period. The combined maximum scale score is 990 points.

Pre-Intervention Testing (Baseline Profile): The official institutional TOEIC scores that the Taiwanese EFL students received from their individual home universities in Taiwan before they departed. The students' individual schools formally certified and sent these baseline measures to create their initial, entry-level English proficiency profiles.

Post-Intervention Testing (Exit Profile): The CPILS (Cebu Practical Academy of Center for Premier International Language Studies), testing center in Cebu, Philippines, officially administered the posttest at the end of the one-month intensive curriculum. To preserve tight concept validity and avoid item-exposure or test-retest bias, this posttest used a parallel, secure version of the institutional ETS TOEIC exam.

Research Procedure

This study looked at a highly structured, brief, intensive language intervention given over a four-week period at the Center for English Language Studies (CELS)

in order to thoroughly assess improvements in English competence. This program's pedagogical architecture was created to integrate Taiwanese EFL learners into a high-density, multi-layered linguistic environment, going beyond conventional rote-memorization methods. An approved institutional pretest administration of the TOEIC Listening and Reading exams in Taiwan was used to establish baseline metrics prior to the start of the intervention. Upon arrival, participants were immediately transitioned into a rigorous academic schedule consisting of four to five concurrent, specialized English language courses. A total of 96 to 120 rigorous teaching hours per student were required each month, with each subject requiring precisely 24 contact hours of direct, in-person instruction.

The intervention's first half was solely devoted to auditory processing techniques. The four subsections of the TOEIC Listening exam—Photographs (Part 1), Question-Response (Part 2), Short Conversations (Part 3), and Short Talks (Part 4)—were methodically taught in Classes 2 through 7 after an introductory orientation. A series of whole-group, small-group, and individual exercises were used in the classroom every day. Recurring mini-tests were used to monitor students' recollection of the material.

On the other hand, text-based processing became the focus of the second part of the curriculum. Classes 10 through 14 concentrated on structural and contextual decoding strategies after Class 9 established a localized baseline using a 100-item Reading Pre-Test. Instruction focused on

context signals for text completion (Part 6), phrase mechanics and syntactic endings for incomplete sentences (Part 5), and skimming/scanning techniques for reading comprehension passages (Part 7).

The concepts of Task-Based Language Teaching (TBLT) served as the methodological foundation for the instruction of these diverse courses (Ellis, 2003). By striking a balance between performance-driven communicative outputs and rich, real-world reading and listening inputs, instructors intentionally operationalized theories of second language learning. Daily lessons allowed students to participate in simulated real-world tasks, work together on group language productions, and explore interactive language functions in place of passive lectures.

The administration of an official, institutional TOEIC Reading and Listening exam at the end of the one-month period marked the conclusion of this intensive pedagogical cycle, enabling direct empirical tracking of how this multifaceted, adaptable intervention affected overall educational efficacy.

IV. ANALYSIS AND RESULT

This section presents the comparative test performances of the respondents between the pretest and posttest as shown in the tables.

Mean and Standard Deviation of the Pretest and Posttest Scores

Table 1: Mean and Standard Deviation of the Pretest and Posttest Scores of the Students

Test	Mean (M)	Standard Deviation (SD)	Verbal Interpretation
Pretest	423.25	152.57	Moderate performance with high variability
Posttest	503.89	173.25	Improved performance, scores still varied

Table 1 presents the descriptive statistics of the pretest and posttest scores of the experimental group. The results show that students obtained a pretest mean of 423.25 (SD = 152.57), indicating moderate prior knowledge with substantial score dispersion among learners. After one month of study as an intervention, the posttest mean increased to 503.89 (SD = 173.25), reflecting an improvement in overall performance. The standard deviation remained relatively high, suggesting that while most students improved, learning gains differed across individuals.

This outcome aligns with prior studies emphasizing the impact of structured language interventions and test-oriented training on learners' communicative competence and performance (Green, 2014; Choi, 2021).

The results can be explained through the lens of Krashen's Input Hypothesis (1982), which posits that language acquisition occurs when learners are exposed to comprehensible input slightly above their current proficiency level. The TOEIC-oriented activities and authentic listening and reading materials used at CELS likely provided such input, allowing students to internalize language structures in meaningful contexts. Additionally, Swain's Output Hypothesis (1985) supports the idea that frequent opportunities to produce language—through speaking, writing, and comprehension exercises. This enables students to test their linguistic hypotheses and refine their accuracy and fluency, which are reflected in their posttest improvements.

Paired t-Test on the Difference Between Pretest and Posttest Scores

Table 2: Paired t-Test on the Difference Between Pretest and Posttest Scores

Comparison	Mean Difference	t-value	df	p-value	Decision on H ₀	Interpretation
Pretest vs Posttest	80.64	11.93	193	< 0.001	Reject H ₀	Significant difference exists

To determine whether the increase in mean scores was statistically significant, a paired samples t-test was conducted, as reflected in Table 2. The analysis yielded a t-value of 11.93 with 193 degrees of freedom and a p-value < 0.001, which is below the 0.05 level of significance. This indicates that the observed improvement from pretest to posttest is statistically significant and not due to random chance. Therefore, the null hypothesis is rejected, and it can be concluded that the intervention had a significant positive effect on students' performance.

The results further resonate with Task-Based Language Teaching (TBLT) principles, which highlight the role of meaningful, goal-oriented activities in language learning (Ellis, 2003). CELS's use of real-life communicative tasks, simulated test items, and integrated skill training likely contributed to the learners' enhanced listening and reading comprehension. This aligns with Lee (2020), who found that targeted TOEIC preparation incorporating authentic tasks and feedback significantly improves learners' test performance and confidence.

Pretest and Posttest CEFR

Table 3: Pretest and Posttest CEFR Linguistic Profile of the Students

CEFR Level	Pretest Count	Percentage	Posttest Count	Percentage
A1	12	6.19%	2	1.03%
A2	169	87.11%	134	69.07%
B1	12	6.19%	52	26.80%
B2	1	0.52%	6	3.09%

The empirical findings reveal an overwhelming upward migration across the Common European Framework of Reference for Languages (CEFR) thresholds, shifting the institutional profile from basic to independent language users. 93.30% of the cohort was highly concentrated in the lower bands at the baseline pre-intervention phase, with 6.19% in the A1 level and 87.11% in the A2 level. This intense focus draws attention to a common phenomenon among East Asian EFL learners, known as the "false-beginner plateau," in which adult students have a sizable passive vocabulary but are stuck at a basic user level because they are unable to actively

navigate structural syntax or handle text constraints under standardized conditions (Richards, 2015).

Following the one-month intensive curriculum, this lower-tier cluster experienced a dramatic decompression: the A1 bracket collapsed to just 1.03%, while the A2 population decreased to 69.07%. At the same time, the upper-intermediate B2 band rose to 3.09% and the independent B1 band surged from 6.19% to 26.80%. From a baseline of 6.71% to an exit measure of 29.89%, this indicates a fourfold increase in independent operational language users (B1 and B2 combined).

Distribution of CEFR Level

Table 4: Distribution of CEFR Level Advancements Post-Intervention

Number of students	Pre test CEFR Level	Post test CEFR Level	Shift
45	A2	B1	A2-B1
3	B1	B2	B1-B2

Table 3 presents the distribution of CEFR level advancements post-intervention. It provides important information about the developmental milestones reached by the Taiwanese EFL cohort. Notably, 45 students (93.75%) successfully completed the transfer from the A2 to the B1 level out of the 48 students who showed an upward trend across global proficiency bands. The A2 level is often described in Second Language Acquisition (SLA) literature as a persistent "proficiency plateau" or false-beginner ceiling, when learners find it difficult to transition from formulaic input to self-regulated language production (Richards, 2015).

The use of Task-Based Language Teaching (TBLT) as an educational framework for high-intensity, short-term courses is strongly supported by this quick increasing trend. Long's (2015) Interaction Hypothesis states that language acquisition happens when students negotiate meaning during communication breakdowns, which makes them aware of interlanguage gaps. Students are rarely encouraged to engage in active linguistic rearrangement in the conventional grammar-translation frameworks characteristic of foreign language contexts.

Additionally, three students attained advanced vertical growth, moving from the B1 level to the B2 level, according to the data. Given the short one-month duration of the intervention, this particular cross-band migration is statistically and pedagogically significant despite being less in number.

V. CONCLUSION

Based on the results of the study, it is imperative to conclude that the lessons, drills, task-based learning activities, practice tests and overall learning experiences provided by the Center for English Language Studies had positively impacted the English Proficiency Level of the EFL students.

Overall, the study also confirms that short-term intensive English programs, when grounded in communicative and task-based pedagogies, can yield measurable gains in EFL learners' proficiency. The students' progression in TOEIC RL scores and CEFR levels demonstrates the effectiveness of CELS's curriculum design, reinforcing the importance of structured input, interactive learning, and performance-based assessment in second language development.

REFERENCES

- [1] Choi, Y. (2021). Washback effects of TOEIC preparation on EFL learners' motivation and strategy use. *Journal of Language Testing and Assessment*, 38(2), 145–162.
- [2] Crystal, D. (2012). *English as a global language* (2nd ed.). Cambridge University Press.
- [3] Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4–58. <https://doi.org/10.1177/1529100612453266>
- [4] Educational Testing Service. (2020). TOEIC® Listening and Reading Test score descriptors and CEFR levels. ETS.
- [5] Educational Testing Service. (2023a). About the TOEIC test. <https://www.ets.org/toEIC>
- [6] Educational Testing Service. (2023b). TOEIC® Listening and Reading Test: Test content and format. ETS. <https://www.ets.org/toEIC>
- [7] Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2019). *Task-Based Language Teaching*. Cambridge University Press. <https://doi.org/10.1017/9781108643689>
- [8] Green, A. (2014). *Exploring language assessment and testing: Language in action*. Routledge.
- [9] Krashen, S. (1982). *Principles and practice in second language acquisition*. Pergamon Press.
- [10] Lee, S. (2020). The impact of TOEIC test scores on employability and career advancement among international students. *Asian EFL Journal*, 24(5), 78–95.
- [11] Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.
- [12] Long, M. H. (2015). *SLA simplified: Critical reactions to recent second language acquisition theory*. Wiley-Blackwell.
- [13] Powers, D. E., & Kim, H. (2019). Developments in the TOEIC program: A retrospective and future perspectives. *ETS Research Report Series*, 2019(1), 1–12.
- [14] Richards, J. C. (2015). *The changing face of language teaching*. Cambridge University Press. <https://doi.org/10.1017/9781139525411>
- [15] Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press.
- [16] Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, 3(4), 207–217. <https://doi.org/10.1111/j.1467-9280.1992.tb00029.x>
- [17] Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House.
- [18] Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in study of language: Studies in honour of H. G. Widdowson* (pp. 125–144). Oxford University Press.
- [19] Swain, M. (2005). The Output Hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–484). Lawrence Erlbaum Associates.
- [20] Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4